Measurement: Sensors 33 (2024) 101118

Contents lists available at ScienceDirect

Measurement: Sensors

journal homepage: www.sciencedirect.com/journal/measurement-sensors

Application of sensor-based speech data mining in E-commerce operations data analysis

Kang Yao

School of Agricultural Information Engineering, Guangxi Vocational and Technical College of Manufacturing Engineering, Nanning, Guangxi, 530105, China

A R T I C L E I N F O

Keywords:
Sensor
Voice data mining
Electronic commerce
Data operation
Data analysis

A B S T R A C T

With the rapid development of the e-commerce industry, enterprises need to effectively manage and analyze various operational data to support decision making and optimize business processes. Sensor technology is widely used in e-commerce environment, which can collect a lot of voice data and extract valuable information through data mining technology. The study presents the current background and trends of the e-commerce industry, with a focus on the importance and challenges of operational data analytics. A sensor-based voice data mining framework is proposed. Sensors are first used to collect voice data and turn it into a digital signal that can be used for analysis. Then the data mining algorithm is used to process and analyze the voice data, and a series of results about e-commerce operation are obtained. Analyze voice data to help companies develop effective marketing strategies, improve product quality, and optimize supply chain management. This study proves the role of sensor-based voice data mining in e-commerce operational data analysis. By harnessing the full potential of voice data, businesses can better understand and meet customer needs, improve competitiveness and performance. This research result has practical significance for the development of e-commerce industry, and provides guidance and reference for the research and practice in related fields.

1. Introduction

With the rapid development of e-commerce, operation data analysis has become a key link for e-commerce enterprises to achieve fine operation and provide personalized services. The continuous progress and popularization of sensor technology has brought new opportunities and challenges to e-commerce operational data analysis. As a device that can sense the environment and collect data, sensors have been widely used in various fields, including smart speakers, mobile devices, etc., providing users with a convenient way to interact. The voice data collected by the sensor has rich information content, and can provide more real and intuitive user experience. Sensors can capture people's voice communication in the e-commerce process, including customer service inquiries, product reviews, and more. With the development of mobile technology, the data on the network is more comprehensive, and the complexity and mobility of the network are increased at the same time, causing the quality of the mobile network to fail to keep up with the scale of the network and the optimized operation of the network [1]. This paper optimizes the impact of VoLTE voice quality in mobile communication networks, and analyzes the data using data collection and voice data mining methods. The research of this paper has two main areas: one is the combination of self-service system and data analysis technology, and the other is the data warehouse technology of building self-service system [2]. The data mining process used in this paper is based on the customer's choice habits (RoadMap) and attributes entered in the self-service voice IVR, using the cluster analysis and correlation analysis methods in voice data mining to achieve customer segmentation [3]. Provide personalized function menus for different types of e-commerce customers, thereby improving the efficiency of e-commerce customers. The premise of data collection in this article is data warehouse technology. E-commerce operation data analysis and management system comes from the actual design of the intern advertising department [4]. The system provides marketing strategy, enterprise analysis, research institute data analysis and data management support for advertisers on the enterprise platform [5]. The system combines internal multi-channel data and provides a variety of artificial intelligence for machine learning, so that users can click on the model to conduct data analysis and help them formulate appropriate marketing strategies [6]. And based on voice data mining, the data is collected and denoised, and then the model is used to analyze and process the customer's behavior and habits, and try to use it to guide the layout of the dynamic voice menu [7]. Use the associated artificial intelligence

E-mail address: dxgdzz@126.com.

https://doi.org/10.1016/j.measen.2024.101118
Received 5 June 2023; Received in revised form 19 February 2024; Accepted 11 March 2024
Available online 18 March 2024

analysis model to conduct data mining on customer behavior, and predict customer behavior after analysis, improve the self-service voice system, from the original static to a dynamic adaptive system [8]. Then combine the data indicators before and after the change to evaluate the scope of improvement. In the context of functional analysis, through case studies and instructions for use, the functions of the system are clarified, and a complete set of system functions are introduced, including sub-modules such as the computer channel data operation analysis system [9]. When analyzing short text data, we need to The required data is analyzed and the structure is created for the artificial intelligence table. Finally, the research and analysis work will be summarized, and the future work will be specifically planned [10].

By mining and analyzing these voice data, we can deeply understand the needs and emotional attitudes of users, and provide more accurate operational decision support and personalized services for enterprises. Voice data collected by sensors also faces some challenges and problems. First of all, the speech data itself has a certain complexity, including waveform shape, frequency domain features, etc., which needs to be processed and analyzed by advanced algorithms and technologies. Secondly, the analysis of speech data needs to take into account the knowledge and methods in multiple fields such as speech recognition, sentiment analysis, and speech recommendation, which involves interdisciplinary cooperation and research. In addition, the privacy protection of voice data is also an important issue, which needs to protect the user's personal privacy while ensuring the effect of data analysis. Therefore, the application of sensor-based voice data mining in e-commerce operation data analysis has important research significance and application value. Through in-depth study of voice data collected by sensors and the development of effective data mining methods and algorithms, e-commerce enterprises can provide more intelligent and personalized operational decision support, improve user experience and service quality, and promote the sustainable development of e-commerce industry.

2. Related work

The literature introduces the optimization based on the data mining model, through the customer model after behavior analysis, research forecasting methods, and optimization [11]. After comparing the improvement degree of the static system and the new system, various methods of data collection, filtering and data analysis are used. It also describes the optimization methods related to the construction of the IVR platform. The literature introduces the process of VoLTE optimization of the hierarchical clustering algorithm for language quality issues, and also analyzes and finds out several direct and fundamental issues that affect voice quality [12]. By using the size of the constraint hierarchical cluster, the root cause is determined based on the size of the constraint. The literature describes the design and implementation of the system [13]. In these steps, we have witnessed the function realization process of the template. For the system realization, it is mainly represented by the class flow diagram and the stack diagram. The literature introduces the requirements of the computer system, mainly the system development environment and the realization of each module [14] These modules include data reporting and platform data monitoring and so on. The literature introduces a Hadoop telecom industry data analysis system framework, which allows organizations to make more informed business decisions through effective data analysis methods. In this case, Flume is used to collect batch behavior data generated by users on the telecommunications provider's website, and then put these in the HDFS file system for data processing on the database [15]. Through Hive statistical analysis of data from different dimensions, and finally using Mahout's collaborative filtering algorithm to recommend products to users.

3. Voice data mining and artificial intelligence technology

3.1. Data mining process

As a kind of equipment that can sense the environment and collect data, sensor is widely used in the field of e-commerce. In e-commerce operations, sensors can capture voice communications between customers and businesses, including customer service inquiries, product reviews, and more. By mining and analyzing these voice data, valuable knowledge and models can be extracted from them to help enterprises make management decisions and predict customer behavior. Voice data collected by sensors can provide rich information, such as customer feedback, emotional attitude, etc., which has important reference value for e-commerce enterprises. Through the processing and analysis of voice data, customers' preferences and needs can be mined to optimize product design, improve customer service and personalized recommendations. By deeply studying the voice data collected by sensors, potential patterns and rules can be found, and more accurate operational decision support can be provided for enterprises. Data mining is a process that uses different ways to extract knowledge and models that are secretly related and meaningful for decision-making in the ocean of data. For example, we take the telecommunication industry as an example. In order to store the information of customer interaction, we establish a large-scale database, and the data knowledge discovery process is to transform the stored data into a process for management decision-making or customer behavior prediction.

Based on the above cognition, it is not difficult to find that the database is a necessary hardware element in the use of data warehouse. Data mining is an analysis method based on them. Databases include different types of structures, but what they have in common is that they can store huge amounts of data and provide the necessary components to support the front-line systems. Data warehouse can realize the functions of data fusion, de-noising, and ensuring the integrity of data. In addition, it has the opportunity to establish a good environment for data mining.

Based on the above methodology, according to our business requirements. In this paper, we often use the data mining steps, can be divided into the following: (1) data selection: select the data consistent with the business needs, ensure that the data is consistent with the goal. (2) Data cleaning: is the so-called de-noising, so that the impact of mining efficiency, or the effectiveness of data cleaning. (3) Data integration: the process of preprocessing the acquired data before analysis. Because different data may be stored in different data sources. We can carry out data fusion according to the needs to pave the way for data mining. (4) Data transformation: used as the basis of data analysis, is one of the important stages of data acquisition. (5) Data mining: using various analysis techniques to realize modeling or patterning. (6) Pattern evaluation: evaluate the relevant knowledge obtained from data mining and confirm its practicability. (7) Knowledge representation: to show the acquired knowledge to relevant personnel, it needs to be clearly visualized. As shown in Fig. 1.

3.2. Principle of voice data processing

If the expected size of mobile network data is too large, the sampling data often contains many noisy, incomplete or even incompatible data, which must be preprocessed with mobile network data object.

$$Z_1 = Z_2 + (Z_3 - Z_2)(t_1 - t_2) / (t_3 - t_2) \tag{1}$$

Data cleaning needs to clean up useless data and some reused data from the original sampling, and then smooth the noise data. There are three ways to deal with the missing value, such as deleting the record, interpolating the data, or we don't need to do any operation.

Data transformation transforms data into a form suitable for mining tasks and algorithms. For different units of data, the size of the data is also different. If the data is not processed, the results of data analysis

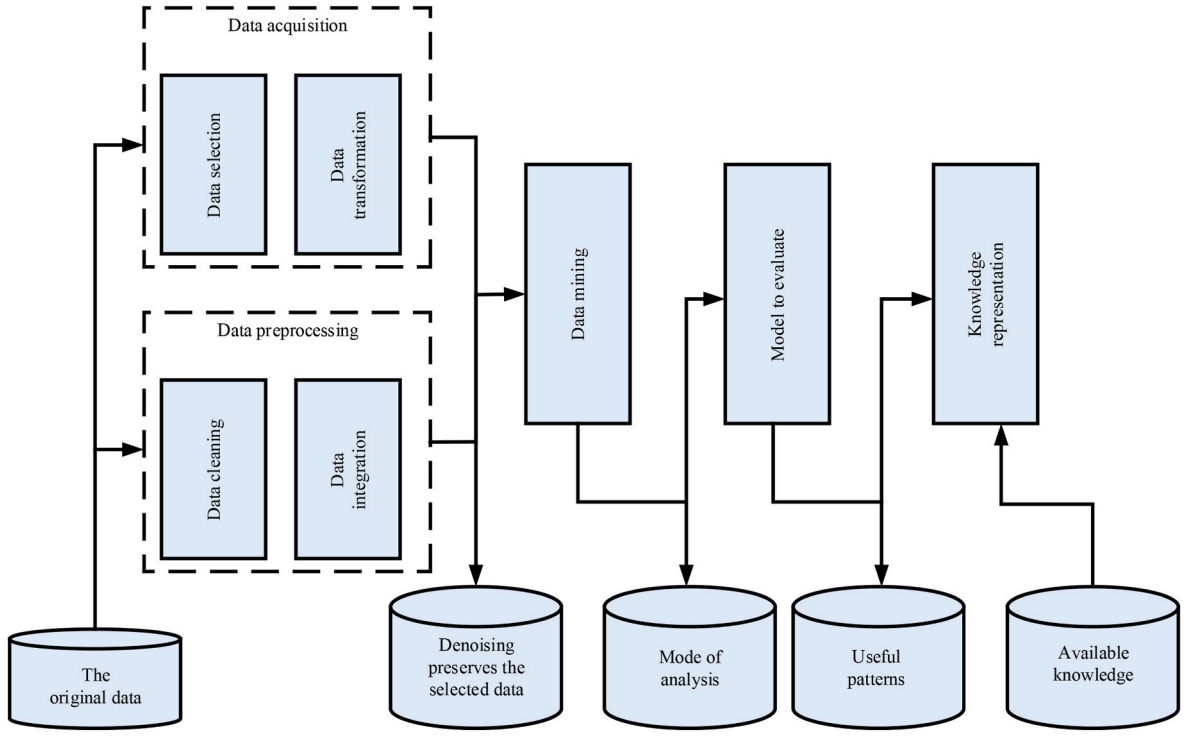

Fig. 1. Schematic diagram of analysis flow.

may be affected. Therefore, standardized scaling is needed.

$$\text{Support } (A \Rightarrow B) = P(A \cup B) \tag{2}$$

$$\text{Confidence } (A \Rightarrow B) = P(B|A) \tag{3}$$

The common distance includes Euclidean distance, which is formula (4):

$$d(i,j) = \sqrt{\left(x_{i1} - x_{j1}\right)^2 + \left(x_{i2} - x_{j2}\right)^2 + \cdots + \left(x_{ip} - x_{jp}\right)^2} \tag{4}$$

Manhattan distance is formula (5):

$$d(\mathrm{i},\mathrm{j}) = |x_{i1} - x_{j1}| + |x_{i2} - x_{j2}| + \cdots + |x_{ip} - x_{jp}| \tag{5}$$

Block distance is equation (6):

$$d(i,j) = \frac{1}{n}\sum_{i=1}^{n} |x_i - \bar{x}| \tag{6}$$

Using cosine similarity measurement for document data, first organize the document data into a document-word matrix format. The formula for calculating the similarity between two documents is equation (7):

$$d(i,j) = \cos(i,j) = \frac{\overrightarrow{i} \cdot \overrightarrow{j}}{\left|\overrightarrow{i}\right| \cdot \left|\overrightarrow{j}\right|} \tag{7}$$

The error square sum SSE is used as the objective function, and the smaller error square sum is selected as the classification result. The continuous attribute SSE calculation formula is equation (8):

$$SSE = \sum_{i=1}^{K} \sum_{x \in E_i} \text{dist}\ (e_i, x)^2 \tag{8}$$

The document data SSE is equation (9):

$$SSE = \sum_{i=1}^{K} \sum_{x \in E_1} \text{cos ine}\ (e_i, x)^2 \tag{9}$$

The calculation formula of cluster center e of cluster E is equation (10):

$$e_i = \frac{1}{n}\sum_{x \in E_i} x \tag{10}$$

In the artificial neural network, the output is completed through the transfer function.

$$y = f\left(\sum_{i=1}^{n} w_i x_i - \theta_i\right) \tag{11}$$

The input layer of the RBF network realizes the nonlinear mapping from x→>R(x), and the output layer realizes the linear mapping from R (x)→yk, as shown in equation (12).

$$y_k = \sum_{i=1}^{P} w_{ki} R_i(x), k = 1, 2, \cdots, q \tag{12}$$

The radial basis function in the RBF artificial neural network is radially symmetric, and the most commonly used is the Gaussian function as shown in equation (13).

$$R_i(x) = \exp - \left[\| x - c_i \|^2 / 2\sigma_i^2\right], i = 1, 2, \cdots, p \tag{13}$$

How to do it: Let the set Dn be the set of corner labels about the point (x, y):

$$P_k(x) = \prod_{i \in B_k} (x - x_i) \Big/ (x_k - x_i) \tag{14}$$

$$L_n(x) = \sum_{j=0}^{n-1} y_j P_j(x) \tag{15}$$

Data transformation is to convert VOLTE voice data into a form suitable for mining tasks and algorithms. This paper mainly uses normalization and standardization in the research of VOLTE voice quality optimization. The feature attribute data that affects the voice quality of VOLTE is scaled equally to eliminate the dimensional difference between the indicators.

$$y_i = x_i - \bar{x} \Big/ s, s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2} \tag{16}$$

Use Euclidean distance to calculate the distance between different categories, as in formula (17):

$$d(i,j) = \sqrt{(x_{i1} - x_{j1})^2 + (x_{i2} - x_{j2})^2 + \cdots + (x_{ip} - x_{jp})^2} \tag{17}$$

In this VoLTE voice quality problem clustering, the number of sample objects is set to n, four direct questions plus twelve bottom questions are used as sample feature indicators, and X is the matrix of samples to be evaluated, then the matrix is equation (18):

$$\mathbf{X} = \begin{bmatrix} x_{1,1} & \cdots & x_{1,16} \\ \vdots & \ddots & \vdots \\ x_{n,1} & \cdots & x_{n,16} \end{bmatrix} \tag{18}$$

Calculate the distance between two combined data. Calculate the distance from the combined data point (A, F) to (B.C), and calculate the average value of the distance between (A, F) and (B, C), as shown in equation (19).

$$d(AF, BC) = (d(A,B) + d(A,C) + d(F,B) + d(F,C))/4 \tag{19}$$

If a random variable X takes the value X = {X_1,X_2, … X_n}, the probability of each is {P_1,P_2, … P_n}, then the entropy of X is defined as equation (20):

$$H(X) = -\sum_{i=1}^{n} p_i \log_2 p_i \tag{20}$$

For the classification system, category C is a variable, and its value is C_1, C_2, … C_n and the probability of each category appearing is equation (21):

$$P(C_i) \approx C_i / |X| \tag{21}$$

And this is the total number of categories. At this time, the entropy of the classification system is expressed as formula (22):

$$H(X) = H(C) = -\sum_{i=1}^{n} p(C_i)\log_2 p(C_i) \tag{22}$$

According to the attribute A, the information entropy of the sample is divided into formula (23):

$$H(X|a) = -\sum_{i=1}^{n}\sum_{j=1}^{m} p(C_i|a = a_j)\log_2 p(C_i|a = a_j) \tag{23}$$

After A divides S, the information gain (Gain) is generated as formula (24):

$$I(X,a) = H(X) - H(X|a) \tag{24}$$

The Gini coefficient calculation formula is formula (25):

$$Gini = 1 - \sum_{k \in I} p_k^2 \tag{25}$$

After being divided into formula (26):

$$\text{Gain} = \sum_{i \in I} p_i \times \text{Gini}_i \tag{26}$$

3.3. Intelligent recognition quality analysis

There are many factors that affect the voice quality of VoLTE networks, including network delay, packet loss rate, voice code frequency and other factors and related conditions, such as cell failures, frequent handovers between regions, eSRVCC switches, etc. By analyzing the recorded VoLTEN language quality data, we can find some factors that affect voice quality.

(1) Speech coding rate analysis

The relative statistical analysis of the MOS value of the speech coding rate and speech quality data is obtained from Fig. 2. The x-axis represents the speech coding rate and the y-axis represents the MOS value. When the speech coding rate decreases under the same bandwidth, the MOS value also decreases. In the same environment, the minimum bandwidth of 12.2 k is the best 3.28. In bandwidth mode, we will get the highest score of 4.14 and 3.14 as the lowest score. The most popular terminal is AMR-wb23.85 kbps. Statistical analysis shows that language coding rate is an important indicator of users' speech perception.

(2) RTP packet loss

The quality of VOLUME language packs may be severely affected. Factors affecting the VoLTE package include error warnings, wireless environment, large-scale voice, transmission, core network, etc. To analyze the dependence of the RTP packet and MOS value of the language quality data, see Fig. 3.

4. Design and application research of e-commerce operation data analysis system based on related technologies

4.1. System requirements analysis

Requirement analysis is a very important part. Requirement analysis includes giving complete, clear and specific development software requirements to determine what tasks the software needs to complete. The combination of user requirements during development is first of all functional and non-functional.

This article analyzes the specific tasks required to organize and store data from the user's perspective. The most important functional requirements of the system are analyzed as follows.

(1) A user name and password are required to enter the system. We also divide the system into two identities, including ordinary users and system administrator users, and verify them according to their database. If the login information does not match or there is an error, An error will be displayed, and if successful, it will jump to the event homepage.
(2) Data storage is a general top-level data collection system for data analysis and a core part of a large-scale data platform. The storage system must meet the requirements for sequential storage and retrieval of structured, unstructured and semi-structured data. Query efficiency and storage security are two important criteria.
(3) The parallel computing system allows the use of MapReduce and Hive operations to process log files in parallel. MapReduce is a programming model that allows parallel computing of large amounts of data. Hive is a Hadoop-based storage tool built on distributed data storage, providing MapReduce operations with

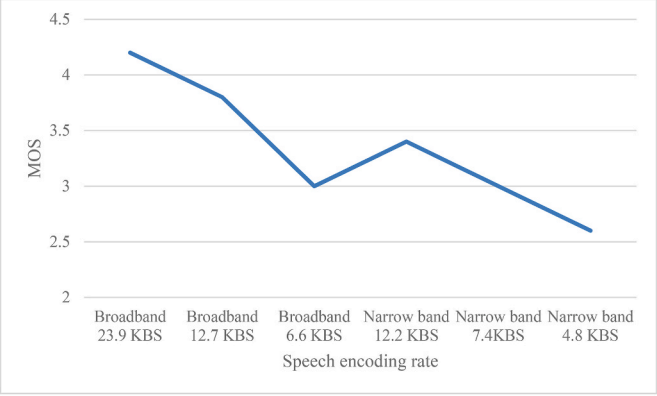

Fig. 2. The relationship between speech coding rate and MOS value.

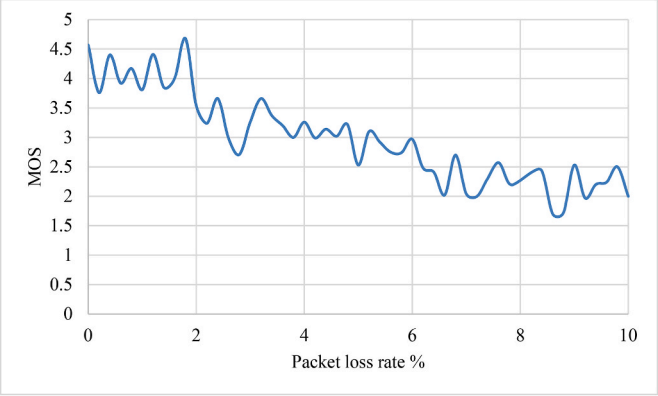

Fig. 3. The relationship between packet loss rate and MOS value.

the ability to implement comprehensive query requests at the lowest level. The system uses MapReduce and Hive to analyze and process log data in real time.

(4) User behavior data analysis includes user data analysis, dialogue analysis, access information analysis and so on. The behavior of counting the number of new users is called user information analysis. A session is the amount of time between a user's login and the complete stop of the system. This interval is called a session and corresponds to the session duration. Browser information analysis is based on user profile analysis, and browser is added as measurement information. Geographic data analysis first analyzes the situation between users and members in different provinces. Research order analysis, including order quantity, charge back quantity and successfully completed orders, and conduct detailed analysis on the big data of users' daily access to predict the general behavior of users. And combine it with marketing strategy to identify the potential problems of current marketing activities, and provide the basis for further improving its marketing strategy.

(5) Commodity evaluation analysis analyzes the complex commodity reviews on e-commerce websites to understand the shopping trend of shoppers.

(6) Recommending advertising products is to recommend more popular, better evaluated and potentially interesting products to users based on user preferences. Provide personalized guidance for customers to improve shopping quality and business value.

(7) Platform management and monitoring can realize the unified online monitoring and management of components in large data cluster, resource logic state setting, timely notification and demonstration of component resource state, and timely discovery and solution of problems through SMS and telephone.

Enterprise analysis system stores data from different departments, and requires daily statistical analysis of business-related data to meet the requirements of system design and implementation. System performance requirements largely depend on product quality, and then affect the requirements of product function. The system performance requirements first focus on reliability, scalability and ease of use.

(1) Reliability may occur due to differences and errors in data collection, so a certain degree of system reliability is needed to run exception information in time when the environment is abnormal, so as to ensure the correct operation of subsequent business logic. At the same time, the system log must be stored in a text file, so that users can query the number.

(2) Scalability: because Hadoop cluster is a distributed cluster, it often exchanges new and old servers, and with the development of the system, it is more necessary to add new services for scalability.

(3) User friendliness to ensure that the system is ready for quick operation.

(4) Other non functional constraints must unify the code style and add necessary comments when writing code.

The overall development flow chart of the system business is as shown in Fig. 4.

4.2. System architecture design

Different from the traditional management system, in the traditional management system, data analysis and processing are based on user needs, and the commercial transaction data generation system is mainly used to support the decision-making of communication operators and other related researchers.

The system consists of four parts: data creation layer, data analysis layer, data layer and infrastructure layer, Hadoop cluster infrastructure, data layer log files and business data analysis process, and finally WeB interface analysis. The following figure shows the overall structure of the system, as shown in Fig. 5.

The data display layer includes login page, master page and data analysis page, which allow system data query and display analysis chart.

Data application layer is used to process the data, including sentiment analysis and recommended products, etc., and will be analyzed in detail according to the user's access information. Recommend products that may be of interest to each user based on their rating, etc.

The analysis layer represents the processing and analysis of relevant data, including data deletion, user behavior analysis, product evaluation analysis and user suggestions. User behavior analysis includes user data, session analysis, browser analysis, geographic data analysis and work analysis.

The data layer is composed of service record files and business data, and most of the service record files are kept structured, which must be cleared and saved to the cache system.

4.3. Database design

The database is the core part of a large number of information systems, and the data table represents the actual content of the database. The database of this system consists of MyS'L and HBase databases. MyS'L database tables are mainly divided into three categories: dimension information tables, fact tables/statistics tables and analysis assistant tables. The measurement information table locates specific measurement information to determine the measurement relationship of the data, such as time. b Platforms, channels, regions, etc. The fact table/statistical analysis result table contains the final analysis data, the foreign key relationship dimension table, and optionally contains all dimension areas as the composite master key. The analysis assistant table contains auxiliary information that can be stored in a relational database (such as MyS'L). b Order information, member information, etc. In the multi-dimensional model, it is mainly the concept level, and the data analysis results are stored as the multi-dimensional model. The tables in the MyS'L database mainly include the following tables:

Browser information Table 1: The main fields include browser number, browser name, and browser version number. The table structure is as follows.

Time information Table 2: The name of the table is dimension date, which stores the information in the time dimension, with dateID as the primary key. The main fields include number ID, year, quarter, month, week, day, etc. The table structure is as follows.

Platform version information Table 3: The name of the table is dimension_platform, which saves the information of the platform version, with platformID as the primary key and foreign key, and cannot be empty. The main fields include platform number, platform name,

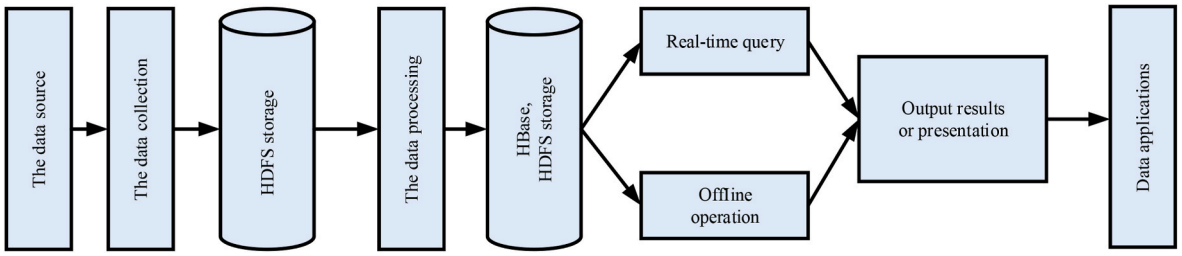

Fig. 4. System business flow chart.

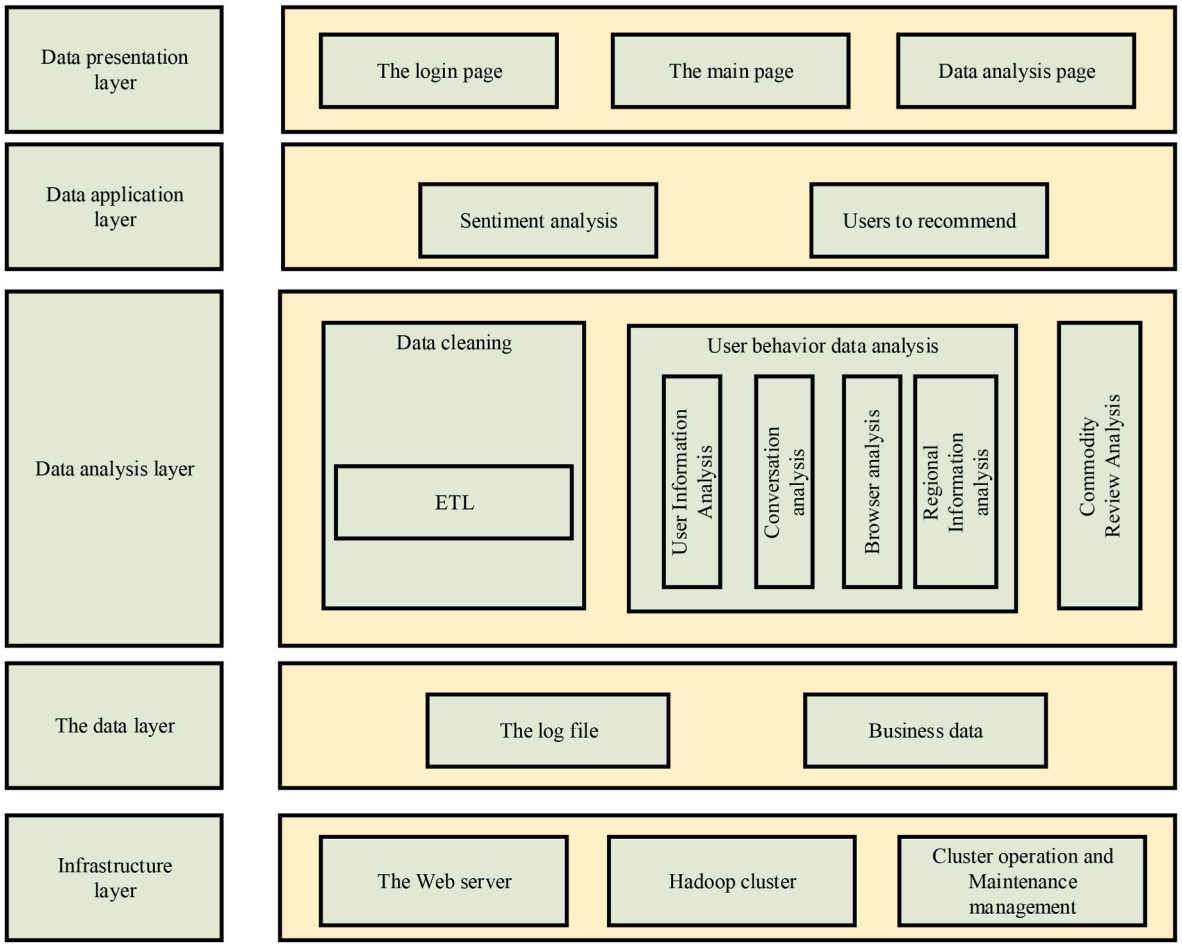

Fig. 5. System overall architecture diagram.

Table 1
Browser information table.

Field name	Field type	Is it empty	Constraint	Remarks
browserID	int (ll)	NOT NULL	Primary key, foreign key	Browser number
browsei˜name	Varchar (45)	NOT NULL		Browser name
browserversion	Varchar (255)	NOT NULL		Browser version number

Table 2
Time information table.

Field name	Field type	Is it empty	Constraint	Remarks
dateID	int (ll)	NOT NULL	Primary key, foreign key	Number ID
year	int (ll)	NULL		Year
season	int (ll)	NULL		Quarterly
month	int (ll)	NULL		Month
week	int (ll)	NULL		Week
day	int (ll)	NULL		H
calendar	date	NULL		Calendar

platform version, and the table structure is as follows.

User browsing information Table 4: The name of the table is rpt_info, which saves the information of users browsing the website, with date as the primary key and foreign key, and cannot be empty. The main fields include time ID, number of unique visitors, page browser, number of logins, number of visitors, average visit time, and number of unique IPs. The table structure is as follows.

User information Table 5: The table is named stats user, which can save information statistics records under each dimension. The main fields are composed of many parts such as user number, time dimension

Table 3
Platform version information table.

Field name	Field type	Is it empty	Constraint	Remarks
platfbrmID	int (ll)	NOT NULL	Primary key, foreign key	Platform number
platfonn name	Varchar (45)	NULL		Platform name
platfbrm version	Varchar (10)	NULL		Platform version

Table 4
User browsing information table.

Field name	Field type	Is it empty	Constraint	Remarks
date	int (ll)	NOT NULL	Primary key, foreign key	Time
uv	int (ll)	NULL		Number of unique visitors
pv	int (ll)	NULL		Page views
loginusers	int (ll)	NULL		Number of people registered
visit users	int (ll)	NULL		Number of visitors
avg time	Varchar (l0)	NULL		Average visit time
ip count string	int (ll)	NULL		Number of independent IP

Table 5
User information form.

Field name	Field type	Is it empty	Constraint	Remarks
userID	int (ll)	NOT NULL	Primary key	User ID
dateID	int (ll)	NOT NULL	Foreign key	Time dimension ID
platfbrmID	int (ll)	NOTNULL	Foreign key	Platform dimension ID
active users	int (ll)	NULL		Number of active users
new install users	int (ll)	NULL		Number of new users
total install users	int (ll)	NULL		Total number of users
sessions	int (ll)	NULL		Number of sessions
sessionslength	int (ll)	NULL		Session length
total members	int (ll)	NULL		Total number of members
activemembers	int (ll)	NULL		Number of active members
new members	int(U)	NULL		Number of new members
created	date	NOT NULL		Creation time

ID, platform dimension ID, number of active users, number of new users, number of sessions Number, session length, etc., but the user's number must not be empty, and other data cannot be empty. They are all key automations, and the time part cannot be empty, as shown in the following table.

4.4. Data processing module

The above is the first step of data processing. We need to use different formats to express the different data types to us, and the collected data should be analyzed in detail. Data integration is one of these technologies, which uses an aggregate format to represent data. A major problem with large data volumes is the exponential growth rate. New technologies are needed to support different types of databases to edit their records, and system data can be collected in two main ways. We can use audit log files to load relevant data into HDFS and collect the data completely. This stage mainly uses JS SDK, Nginx, Flume configuration. However, first you must know which events to find.

Launch event is a type of event that a user (visitor) applies to a website for the first time, and is mainly used to calculate new users.

Page View event is the main event type on the PC page. The first step is to describe the website information of user visits and computer tasks in detail. Used to use jf to record the triggering behavior of specific events/activities to calculate the conversion rate of each activity's active users and each link that can be accessed.

Nginx is a powerful reverse proxy server, which is often used with Tomcat to take advantage of dynamic resource partitioning, load balancing, etc. Nginx requires less memory, is faster, and can be easily expanded. For static resources, Nginx can provide high-speed access services for them, and it is quite convenient, but at most it can only support request information from one server. Nginx usually acts as an access server or traffic server for static resources. Flume, as a reliable and useable system, has a great effect. It can not only transmit, but also realize large-scale aggregation. You can configure different types of data to the sender in the data collection protocol system. In addition, Flume can also simplify data processing a lot, write data and send it to different recipients. Organizations often use weather vanes to monitor ports or files, extract recently added content to other locations, or monitor directories, and retrieve the files it contains on the HDFS platform in real time.

The World Wide Web is a growing library of millions of volumes without a centralized management system. Just like the fish in the haystack. Different search engines solve this problem by indexing the complete Imemet content. To this end, search engines use an automated program, a web-based scanner. The most important task in the network is to obtain information and corresponding efficiency. To help search or archive, this can also be achieved through a network scanner. The web scanner can access all the links automatically, and these links are further indexed. However, the use of network scanners is not limited to search engines. It can also be used to cancel the Internet and process useless emails. In addition, it can also identify some unauthorized content, such as comprehensive network content, multimedia content, and so on. The network scanner structure used in this article is the main network scanner infrastructure of Python. It is mainly composed of scanner, URL manager, HTML loader, HTML parser and five most important storage modules.

The crawler scheduler is mainly responsible for coordinating the work of the other four modules. The URL manager is responsible for maintaining a collection of URLs, including extracted and unextracted URLs, and can also provide an interface for new links. The HTML downloader can accept links that have not yet been verified, and then download the website's information. The HTML parser analyzes the URL references and trusted data of the collected WeB information, passes it to the URL manager and stores the data.

5. Conclusion

In the field of e-commerce, sensor-based voice data mining has the potential to be widely used in operational data analysis. By using sensor technology to collect users' voice data on e-commerce platforms, it can provide valuable information and insights that can help businesses optimize operational strategies, improve user experience and increase profitability. Sensor-based voice data mining can help businesses understand users' purchasing preferences and behavior patterns. Through the analysis of voice data, it can identify the key decision-making factors of users in the purchase process, evaluate the satisfaction of goods or services, and purchase intention. This data can be used to optimize product recommendation algorithms, personalization, and marketing strategies to improve sales and customer loyalty. Voice data mining can help companies identify and solve problems that users encounter when

using e-commerce platforms. By analyzing users' voice feedback and complaints, deficiencies in technical faults, logistics issues, customer service quality, etc. can be identified and resolved in a timely manner. This ability to solve problems in a timely manner helps to improve user satisfaction and increase user trust and loyalty to the e-commerce platform. Sensor-based voice data mining can also be applied to market competition analysis and business intelligence gathering. Through the monitoring and analysis of voice data, information such as competitors' marketing activities, product releases and promotion strategies can be obtained, thus helping enterprises to make more accurate market decisions and strategic planning. This collection of business intelligence can provide insight into market trends, consumer needs and the competitive environment, supporting companies to gain a competitive advantage in the highly competitive e-commerce space. Therefore, sensor-based voice data mining has the potential to be widely used in e-commerce operational data analysis. By using voice data mining and analysis, enterprises can obtain important information about user purchase preferences, behavior patterns, and feedback on problems, so as to optimize operational strategies, improve user experience and increase profits. However, it should be noted that privacy protection and data security still need to be valued and guaranteed to ensure the lawful use and protection of user data.

Ethics approval

Not applicable.

Funding

There is no funding

Informed consent

Not applicable.

CRediT authorship contribution statement

Kang Yao: Conceptualization, Data curation, Formal analysis, Writing – original draft.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

References

[1] H. Lu, Y. Zhang, Y. Li, C. Jiang, H. Abbas, User-oriented virtual mobile network resource management for vehicle communications, IEEE Trans. Intell. Transport. Syst. 22 (6) (2020) 3521–3532.
[2] S.A. Lonkar, K.V. Reddy, Analysis of audio and video quality of voice over LTE (VoLTE) call, Int. J. Inf. Technol. 14 (4) (2022) 1981–1994.
[3] A. Rotondo, F. Quilligan, Evolution paths for knowledge discovery and data mining process models, SN Computer Science 1 (2) (2020) 109.
[4] R.J. Kauffman, J. Srivastava, J. Vayghan, Business and data analytics: new innovations in the management of e-commerce, Electron. Commer. Res. Appl. 11 (2) (2012) 85.
[5] S.S. Alrumiah, M. Hadwan, Implementing big data analytics in e-commerce: vendor and customer view, IEEE Access 9 (2021) 37281–37286.
[6] Q. Zhou, Z. Zhang, Y. Wang, Research on safety management system optimization of B2C e-commerce intelligent logistics information system based on data cube, J. Intell. Fuzzy Syst. 38 (2) (2020) 1585–1592.
[7] A. Toleuuly, B. Yessengeldin, Z. Khussainova, A. Yessengeldina, A. Zhanseitov, S. Jumabaeva, Features of e-commerce risk management in modern conditions, Acad. Strat. Manag. J. 19 (1) (2020) 1–6.
[8] Y. Tang, Y. Huang, H. Wang, C.Q. Wang, Guo, W. Yao, Framework for artificial intelligence analysis in large-scale power grids based on digital simulation, CSEE Journal of Power and Energy Systems 4 (4) (2018) 459–468.
[9] H. Chen, S. Chen, J. Zhao, Integrated design of financial self-service terminal based on artificial intelligence voice interaction, Front. Psychol. 13 (2022) 850092.
[10] Y. Vakulenko, P. Oghazi, D. Hellström, Innovative framework for self-service kiosks: integrating customer value knowledge, Journal of Innovation & Knowledge 4 (4) (2019) 262–268.
[11] S.N. Mohanty, E.L. Lydia, M. Elhoseny, M.M.G. Al Otaibi, K. Shankar, Deep learning with LSTM based distributed data mining model for energy efficient wireless sensor networks, Physical Communication 40 (2020) 101097.
[12] J. Kreiman, Y. Lee, M. Garellek, R. Samlan, B.R. Gerratt, Validating a psychoacoustic model of voice quality, J. Acoust. Soc. Am. 149 (1) (2021) 457–465.
[13] B.R. Rocha, M. Behlau, The influence of sleep disorders on voice quality, J. Voice 32 (6) (2018) 771-e1.
[14] L. Liu, S. Lu, R. Zhong, B. Wu, Y. Yao, Q. Zhang, W. Shi, Computing systems for autonomous driving: state of the art and challenges, IEEE Internet Things J. 8 (8) (2020) 6469–6486.
[15] H. Zahid, T. Mahmood, A. Morshed, T. Sellis, Big data analytics in telecommunications: literature review and architecture recommendations, IEEE/CAA Journal of Automatica Sinica 7 (1) (2019) 18–38.